CHAPTER 188

Massachusetts
Educational
Assessment
Program

November 1989

On Their Own:

Student Response to Open-Ended Tests in Science

MASSACHUSETTS DEPARTMENT OF EDUCATION

MASSACHUSETTS BOARD OF EDUCATION

Mr. James F. Crain, *Chairperson*, Gloucester
Dr. Raquel Bauman, *Vice Chairperson*, Holden

Mr. John J. Gould, Waltham
Mr. James R. Grande, Hanover
Mr. Joseph E. Killory, Sandwich
Mr. Sol Kolack, Newton
Mrs. Anne S. Larkin, Bedford
Mr. Edward T. Moore, Jr., Bellingham
Ms. Theodora A. Sylvester, Springfield
Mrs. Frances M. Turner, South Egremont
Dr. Joan Wallace-Benjamin, Boston
Mrs. Mary C. Wright, Falmouth

Harold Raynolds, Jr., *Commissioner of Education, Secretary of the Board of Education*

Dr. Franklyn Jenifer, *ex officio Board Member, Chancellor, Board of Regents of Higher Education*

OFFICE OF PLANNING, RESEARCH AND EVALUATION

Allan S. Hartman, Director of Research

M. Elizabeth Badger
Beverly Miyares
Brenda Thomas
Mary Robertson

On Their Own: Student Response to Open-Ended Tests in Science

Elizabeth Badger

Brenda Thomas

Massachusetts Educational Assessment

Massachusetts Department of Education

1989

The Commonwealth of Massachusetts
Department of Education

1385 Hancock Street, Quincy, Massachusetts 02169-5183

November 8 1989

Dear Educator:

The four booklets in this series discuss the reading, mathematics, science and social studies results of the 1988 Massachusetts Educational Assessment Program. They represent one of the many efforts of the Department of Education to help schools carry out their educational mission more effectively. In this case, they provide models for student evaluation within the classroom, as well as describing students' progress in understanding.

The title of this series, **On Their Own**, suggests an important aim of education: the ability of students to act as independent, rational thinkers. The questions described in these booklets demand that ability. They demand active intelligence as students are required to relate what they know to new and challenging situations.

In addition to describing students' understanding, these booklets carry a message about the evaluation that goes on in the classrooms. The message is that the short objective tests of facts or procedures, standard fare in most classrooms, are too slight a vehicle to convey the true purpose of evaluation.

In the first place, effective student evaluation is an important component of effective teaching. Research has described the complex thinking that underlies students' errors and misconceptions. Unless teachers take the time to discover for themselves how students understand a subject, they will be unable to adjust their teaching in appropriate ways. This kind of evaluation, involving student discussion and explanation, should be a continuous and constant part of every classroom.

Secondly, evaluation can, and does, affect students' learning. Not only does it signal for the student the content areas that teachers consider important, it gives a message about the kind of thinking that is considered valuable. When testing is limited to short objective questions, requiring a single answer, the message given is that facts are what really count. When questions encourage students to think, to grapple with the material and to consolidate different aspects of learning, the message is much different. Such questions indicate to students that it is the quality of thought that is important, not the correctness of the answer itself. The possibility of different answers opens the door for discussion, argumentation, and intellectual excitement in our schools. This is the message that we want to convey to our students.

We hope that you will study the material included in this series and incorporate the ideas presented in your own classrooms.

Sincerely yours,

Harold Raynolds Jr.

Harold Raynolds Jr.
Commissioner of Education

Acknowledgments

This report would not have been possible without a major contribution from members of the Science Advisory Committee. These teachers and science coordinators analyzed the responses for each question, read and scored the scripts, and interpreted the results with reference to both student achievement and school instruction. It was a major project, which they accomplished with competency, efficiency, grace, and goodwill. If you find this book at all useful, it is they who should be thanked.

Members of the Science Advisory Committee who contributed to making this booklet possible are:

Mary Corcoran	**Winthrop Public Schools**
Kathleen Donnellan	**Springfield Public Schools**
Thomas Heyman	**Weymouth Public Schools**
F. Paul Quatromoni	**Dedham High School**
Mary Ann Sudolnik	**Turkey Hill Middle School, Lunenburg**
Diane Paradis	**Assabet Valley Regional School**
Henry Kaiser	**Pittsfield Public Schools**
Angela Johnson	**Cambridge Rindge and Latin**

In addition, we would like to thank Allan Hartman of the Office of Planning, Research and Evaluation for his helpful comments, and we are particularly grateful to Stuart Kahl of Advanced Systems for his work on the development and analysis of the test questions.

Elizabeth Badger
Brenda Thomas

Table of Contents

Foreword

Senta A. Raizen
Director, National Center for Improving Science Education

Dissatisfaction has been growing with common methods of assessing students' learning in science, most often through multiple-choice tests. The dissatisfaction is due to the mismatch between the goals of science education and what multiple-choice or other short-answer questions, each with one preordained "right" answer, can tell us about students' knowledge and competence in science. These sorts of questions are useful in assessing recall of factual information and some deductive reasoning skills, but they don't address such important goals as learning to understand how scientists think and why they think what they think, applying science knowledge to individual and societal problems, designing ways of addressing unfamiliar or puzzling situations, and constructing solutions. At any level of science knowledge, we hope that our students are able to exercise these complex skills, appropriate to their age and development.

The Massachusetts Department of Education, therefore, is to be commended for going beyond the usual testing format and trying to probe these critically important but more elusive learnings. The results clearly indicate that there is a wide gap between recognizing the correct formulaic answer when it is provided as one of several choices and either applying that knowledge or demonstrating an understanding of its meaning, as the "Spot Remover" and "Tree" examples demonstrate. One wonders how much greater the gap would be if students were confronted with a situation that was not spelled out for them, in which they would have to define the problem(s) to be investigated, decide what data to collect and how to analyze these data, formulate some conclusions, and then decide how to test these conclusions further.

Another matter of interest is the ability of students to use equipment and apparatus in their science investigations. After all, the uniqueness of science

is in the use of the senses – as extended through physical tools that greatly enhance human powers of observation – to gather evidence on how the world around us works. There has been some experimentation with hands-on assessment tasks both in Great Britain and in this country, including Massachusetts, that has demonstrated their feasibility, even for sizable samples of students. At this point, Massachusetts plans to continue this approach as part of its assessment program.

Even the more limited attempt, recorded in this report, to expand written assessment of science learnings beyond those addressed by traditional tests deserves close attention. The results themselves are illuminating, but perhaps even more important is the message conveyed to teachers, administrators, and parents regarding the importance of previously unaddressed goals of science education. Also, one hopes that innovations in state assessment will be reflected in improved tests given by teachers in individual classrooms.

Readers should approach the assessment examples and the results before them in the spirit of science – as an experiment. Do you agree with the interpretations and conclusions? Are you satisfied with the approach? How might one be able to find out more about the science knowledge and competencies acquired by Massachusetts students? Does the level of student achievement in science, as portrayed in this assessment, meet your expectations? If not, what do the findings say about next steps in the classroom, at the district level, at the state level?

If this report spurs readers to ask themselves questions about science instruction, what students take from their instruction, and how their knowledge and their ability to use this knowledge ought to be assessed, then the experiment will have been worthwhile. And if readers will take action, at the state, district, or local school level, to improve students' science learning and how it is assessed, then the Massachusetts Department of Education surely will continue its experimentation to provide richer and educationally more meaningful information on what students in the Commonwealth know and can do in science.

Introduction

In the spring of 1988, the Massachusetts Department of Education administered its second biennial assessment, testing fourth, eighth, and twelfth grade students in four content areas – reading, mathematics, science, and social studies. Although the large majority of the over 3000 items were given in multiple-choice format, some of the items were open-ended, requiring students to answer in written form. These open-ended questions appeared in one form of the tests at each grade level. Consequently, one-twelfth of the fourth grade students, one-sixteenth of the eighth grade students, and one-twentieth of the twelfth grade students received a test form that contained some open-ended questions. However, they did not all receive the same questions. The ten or so questions in each subject area were distributed in such a way as to produce a sufficiently large number of responses to each question to report reliably. It should be understood, however, that with this limited number of questions we did not attempt to cover all methods, principles, and knowledge bases that each subject requires. This was a sampling rather than a complete assessment of competence in any one subject.

There are three reasons for our decision to include open-ended questions in our assessment of student performance. The first is the value of the information obtained. While multiple-choice items are efficient, easy to score and objective, they are a weak measure of how students actually think. Neither can they measure students' ability to generate solutions (rather than pick from a list already prepared for them). Nor do they test the students' approach to those ill-structured problems that are most familiar in everyday life. Including open-ended questions on the assessment results in a more valid estimation of student achievement than we would have obtained had we limited assessment to multiple-choice items.

Our second purpose in including these open-ended questions was our belief in their intrinsic value: they call for the kind of thinking that education is all about. Too often educators pay lip service to the need for active learning but teach and test students in ways that demand passivity. By their actions, schools say to students, "We are not interested in *your* response; we are only interested in the *correct* response." This report of the open-ended testing

shows how students respond when they are challenged to define a problem as well as deal with it.

Finally, we hope that such testing will act as a model for classroom testing. We reproduce the questions themselves here, as well as report on state-wide results, so that teachers can try them out in their own classes and, if they desire, compare the results they obtain with the state norm. Our intent is to show that this type of testing yields important information about students' understanding of concepts and procedures, their ability to apply their learning to new situations, and their need for further instruction.

Although the results in each subject area are treated in separate booklets, the underlying thought processes which we report on are similar. They reflect an approach to thinking that stresses engagement and critical evaluation. Beyond this, however, we look at how students function in various learning contexts which require the understanding of specific concepts.

The common theme across all the subjects tested is that learning is a process of constructing meaning, of restructuring what we know and believe to accommodate new information and ideas. It is not passive, nor is it easy. It involves reflection, questioning, comparing, and adapting – the kinds of intellectual activities that we sum up as "critical thinking."

The link between critical thinking and learning is possibly most apparent within the area of reading, where experts point to the close connection between reading comprehension (the ability to infer) and active analysis. (This is discussed in the accompanying report, **On Their Own: Student Response to Open-Ended Tests in Reading.**) In a more content-ladened subject, such as science, it may be less apparent. However, science educators increasingly call for more active involvement of students at all levels. In doing so, they point to the major goals of science education:

> the ability to use scientific thinking (i.e., critical thinking) to analyze and evaluate;
>
> an understanding of what is meant by the scientific process (hypothesis testing, theory making) and the basically tentative nature of science;
>
> an understanding of some of the major theories and constructs that scientists have proposed as explanations for the world.

This emphasis on broadly based skills and understandings that should be possessed by *all* students is essentially different from past concentration on more specific skills for the scientifically gifted, and it holds at least three important implications for education.

- **Students must be allowed the opportunity to practice thinking and acting like scientists.**

Scientific inquiry is not merely a set of rules or procedures. It is the natural extension of an attitude, a way of thinking about information in the world. In order to understand this, students should be asked to explore, hypothesize, collect information, organize and interpret data. They should be able to evaluate the investigations of others.

This approach to science education reflects the evolving nature of science itself and should serve students well in later years. Daily life for many graduates is far removed from the systematic exposure to science provided in a school setting. Certain facts, theories, and instruments that they will encounter in school may well be superseded by further developments. At the same time that they are becoming "rusty," they will be bombarded by competing claims to "scientific" authority in the course of media blitzes and marketing campaigns, and as citizens they will confront issues that revolve in perhaps unforeseen ways around technology, environmental problems, public health policy, and the like. The best preparation is a sense of what a scientific approach requires, regardless of the particular context.

- **Students' ideas should be listened to and respected.**

From the moment that children begin to interact with the world, they begin to build theories and explanations for how things function. By the time they reach the more formal environment of the classroom, they already possess both a mass of information about the world and beliefs about how and why it works as it does. Some of these beliefs are similar to those meanings generally accepted by the scientific community. Some are not. However, the demand for personal meaning gives great force to children's own explanations. These explanations are not easily abandoned; nor are they replaced on the basis of teacher instruction. Teachers need to know how their students make sense of the world and, if necessary, stimulate them to reexamine their theories. The point is well made in a recent article by J. Minstrell:

> The act of instruction can be viewed as helping the student unravel individual strands of belief, label them, and then weave them into a fabric of more complete understanding. An important point is that later understanding can be constructed, to a considerable extent, from earlier beliefs. Sometimes new strands of belief are introduced, but rarely is an earlier belief pulled out and replaced. Rather than denying the relevancy of a belief, teachers might do better by helping students

differentiate their present ideas from and integrate them into conceptual beliefs more like those of scientists.[1]

- **Learning should take place in the context of real tasks.**

There are two justifications for this recommendation. One is based on cognitive science, which has investigated the role that information plays in the thinking of experts. Here, the consensus appears to be that it is not the amount of information *per se* that distinguishes the expert from the novice, but the extent to which that information is organized into a meaningful structure.[2] What concerns teachers is how to develop such awareness in their students.

It is doubtful that well-defined, isolated, made-to-order tasks promote students' understanding of the relationship among ideas. Such appreciation is more likely to result from experience with complex tasks that more clearly reflect the ambiguities found in real life.

The second justification for real-life context is motivational: the extent to which students recognize school science as a relevant part of their lives. It is ironic that, in a "scientific age," when newspapers and television refer frequently to scientific achievements and controversies, students should view school science as irrelevant and uninteresting. And yet, study after study reports that this is the case. Children's initial wonder and fascination with learning about how the world works becomes increasingly dampened the longer they remain in school. By twelfth grade, only 30 percent of Massachusetts students found science very useful in understanding the world about them, and only 12 percent found it very useful for getting along in everyday life.[3] Something has to be done to change this perception.

The set of open-ended questions which we presented on the assessment attempted to reflect these concerns. We asked students two types of questions – one that required them to act like scientists, using different aspects

1 Minstrell, J. "Teaching Science for Understanding," in L. Resnick et al. (eds.), *Toward the Thinking Curriculum: Current Cognitive Research* (ASCD Yearbook, 1989), p. 130.

2 Bransford, J., and Vye, N. "A Perspective of Cognitive Research and Its Implications for Instruction," in Resnick, *ibid.*, pp. 173-205.

3 *The Massachusetts Educational Assessment Program: 1988 Statewide Summary.* Report issued by the Massachusetts Department of Education, 1988.

of the scientific approach to answer questions; the other that investigated their understanding of some fundamental concepts in the field of life and physical sciences. We attempted to place all questions in the context of real problems. Our purpose is not only to report on how well students performed in response to these tasks, but to illustrate the kinds of questions that might form the basis for both instruction and evaluation within the class. When practical, each of the questions is also accompanied in the Appendix by a chart giving the percentages of responses throughout the sample. These are presented for the convenience of teachers who might wish to compare the performance of their own students with those across the state.

Notes

A Question of Answers

Scientific Inquiry

There is a large difference between knowing rules and actually using a scientific attitude to explore situations. Generally, students perform well when the problems are well defined and well structured, as they are in multiple-choice exercises. Open-ended questions by contrast allowed us to explore how well students are able to apply their knowledge in a less structured context which more closely resembles that of everyday life.

The four open-ended questions that we used to test the sense of what scientific inquiry entails covered a number of different aspects of what students might be expected to know.

- The one called **Spot Remover** was designed to test students' recognition of the importance of controlling variables in an experiment. At the fourth grade, children were merely asked to recognize the important factors that should be controlled. At the older levels, they were expected to recognize the need for control, as well as to identify those factors and to provide an experimental design.
- The exercise called **Homework** was given only to twelfth graders. Again, it dealt with the control of variables in an experiment; however, it set a more analytical task than **Spot Remover** in that students were asked to critique an experimental design. This required their identifying the experimental variable, as well as recognizing the possible effects of other variables on the result.
- **Wanda** is an exercise that tested both concepts and procedures at the eighth and twelfth grade levels. Requiring students to generate the essential questions that would test a set of hypotheses, it measured students' understandings of the meaning of *hypothesis* and the need for an operational definition of a relative term. By asking students to describe how they would reduce and display the results, it investigated how well they understood the

limitations of different types of variables and how well they could use graphical techniques.

- **Brothers and Sisters**, the final exercise under the Scientific Inquiry category, explored students' ability to interpret data presented in tabular form. Grade 8 students were told how to proceed with the interpretation, while Grade 12 students were given more open-ended directions to "use the data."

Thus, the four open-ended questions that investigated students' scientific inquiry skills represent a continuum from a completely unstructured situation in which students are asked to identify relevant variables to one in which the variables are defined and student are required to organize data and draw conclusions. What happened when they encountered these questions is described in detail below.

Spot Remover: Grade Four

> Chris wants to find out which of two spot removers is better. He decides to try out each of the spot removers on some stains. He tests Spot Remover A first. What are some things that should be the same when he tests Spot Remover B?

The simple exercise shown here concerning the curiosity of a boy named Chris about the relative effectiveness of two kinds of spot remover examined Grade 4 students' understanding of experimental design. The question directed them to consider elements of the situation that should be kept the same. Answers were judged in terms of reasonableness and the number of elements cited. A child who wrote "the stain should be the same" or "the material should be the same" was deemed to be on the right track.

Apparently this is an unfamiliar and daunting task to the majority of fourth graders, although it may be for lack of practice rather than lack of cognitive development. Some 15 percent of the respondents left the lines blank, and an additional 42 percent failed to name any components relevant to an experiment – such as type of fabric, amount of spot remover applied, and time allotted for drying. Confused by the latitude of the question and the lack of structure, many students misinterpreted "things that should be the same." They thought the phrase referred to the spot removers and not to the elements of a controlled comparison. Thus a number of responses that we classified as inappropriate discussed possible similarities of the substances in question. Two typical answers:

> Spot Remover B will probably be the same color, or smell the same or have some of the same stuff in it, or it might cost the same or be from the same store.
>
> It will take a stain out but not very good as the other. I don't think it's the same.

Nevertheless 30 percent of the respondents did identify one or two elements appropriate for controlling in the experiment, and a final 13 percent came through with flying colors by naming three or more such elements. Students

who responded correctly tended to write detailed answers, as in the following sample:

> Use the same kind of stains like ketchup and tomato sauce, soup, grape jelly, raspberry jelly, jam. Use the same kind of cloths and the same colors.

Spot Remover: Grade Eight and Grade Twelve

> A person wants to determine which of two spot removers is more effective. Describe in detail an experiment the person might perform in order to find out which spot remover is better for removing stains from fabrics.

For Grade 8 and Grade 12 students, the wording of the spot removal question was changed to put them on notice that their sense of valid experimentation was being tested. The need for control was not mentioned. They had to recognize the requirements of the task for themselves.

Results obtained from similar multiple-choice questions indicate that increasing age – and probably, for twelfth graders, experience in laboratory courses – leads to greater awareness of the complexity of experimental design and the need to control variables. Students often perform well in this area. The results illustrated by the following multiple-choice question are fairly typical.

Which of the following is essential in an experiment?

Gr 8	Gr 12		
13%	5%		Making sure measurements can be made quickly
55	81	*	Controlling all important variables
7	2		Using new equipment
24	11		Having at least two persons doing the experiment

The purpose of the open-ended task was to put this knowledge to a further test. Given the large number of students, particularly at the twelfth grade, who acknowledged the necessity of controlling variables in an experiment, the open-ended question assessed their ability to recognize and apply that knowledge in an everyday situation. In this less structured context, they found the task difficult.

Approximately a quarter of students at both levels did not recognize the need for controls at all. Another quarter of eighth graders and somewhat fewer twelfth graders mentioned only stain and fabric as relevant factors in comparing the two products. Of those students who described a satisfactory experiment, the difference in age and experience emerged in the older students'

recognition of the need for replication. Forty-two percent of the twelfth graders described an experiment in which the initial procedures were replicated with factors controlled differently (e.g., different stains, different fabrics, different duration). Only 8 percent of eighth graders were this thorough. Most gave a general answer, with no reference to specific procedures. (See Appendix for classification of results.)

Homework: Grade Twelve

A teacher wanted to know if homework really made that much difference in how much science students learn. He had two classes of students of equal ability. During one three-week unit of instruction, he assigned no homework to one of the classes. Instead, he shortened their lab periods and used that lab time to have those students complete extra worksheets. The other class had full-length lab periods and regular homework. At the end of the unit, the two classes performed equally well on a test, so the teacher concluded homework does not make a difference. Is the teacher's conclusion a good one, or could the experiment have been improved? Explain your answer.

This open-ended exercise that invited respondents to critique an experiment with homework was administered to Grade 12 students. The most appropriate answer here would have been that the hypothetical teacher in the exercise should have given both classes a full lab period and contrasted homework with no-homework (rather than worksheets that served as an equivalent of homework).

Some 20 percent of the students stuck to the main issue and expressed it adequately. In the words of one of them:

> I think the teacher's conclusion was incorrect. Even though the first group did not have assignments to be completed at home, they did get extra practice with the added worksheets. Without these, they probably would have performed differently on the test.

A further 15 percent of the students grasped the main point but mixed in less convincing concerns.

> His test sample is too small to get any accurate results to begin with. There are also too many substitute factors such as worksheets and shortened lab time which make the classes unequal. To improve the test the teacher needs more tests with a greater number of students when all conditions but the doing of homework are the same.

For the remaining two-thirds of the students, the task set here was apparently beyond comprehension. Fully 18 percent of them gave irrelevant answers or left the lines blank, and a further 25 percent used the exercise as a forum for expressing their opinions about homework. "It's a good [experiment] because homework makes no difference. I never got a lot out of homework," one of them wrote.

Almost another quarter of the students stated a conclusion as called for by the exercise but offered no reasons or gave poor reasons. "The conclusion was a good one, but I feel that it all depends on how much students are willing to study" fell into this category. "There are a lot of factors in his experiment for he only used classes of equal ability. What about the classes that doesn't do that good on work?" worried one student. (See Appendix for complete results.)

That only 36 percent of the students were able to successfully critique this experiment casts doubt on student understanding of the issue of controlling variables in an experiment. As cited previously, over 80 percent of twelfth graders were able to identify the need for controls when the question was asked as a factual item. However, when asked to apply that knowledge, they did not recognize its relevance.

Wanda: Grades Eight and Twelve

1. Wanda wants to test the following hypotheses: (1) girls who play a musical instrument are not as good in math as girls who do not play a musical instrument and (2) the opposite is true for boys. She will ask many students to complete a questionnaire to get the information she needs. What are the questions that Wanda must include in the questionnaire?

2. Explain how Wanda might analyze and display her results. (For example, how could she set up a table or graph to help her test her hypotheses?)

Told about a project that an imaginary student named Wanda intended to set up, Grade 8 and Grade 12 students were challenged by two open-ended questions to indicate what should be done to ensure a valid investigation.

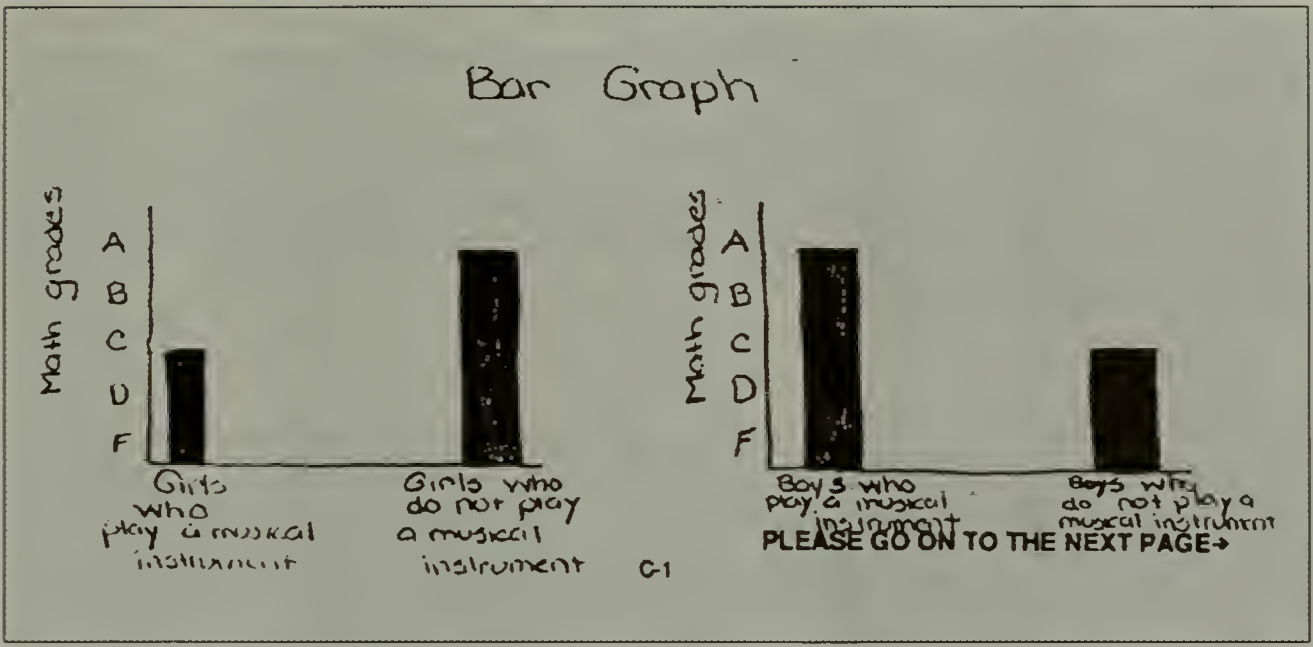
Although the wording might vary, the first order of business for Wanda was to collect certain information from everyone she approached. *Are you male or female? What is your grade in math? Do you play a musical instrument?* Some 42 percent of our eighth graders and 58 percent of our twelfth graders generated these essential questions. In addition, a number of students who neglected to mention male/female at the questionnaire stage did in fact provide a graph for girls and a graph for boys at the display stage. In contrast, almost a fourth of eighth graders (but only 5 percent of twelfth graders) responded irrelevantly, indicating a misunderstanding of the term *hypothesis*. Many seemingly accepted Wanda's hypotheses as established facts, not propositions to be tested, and interpreted the task as one of discovering *why*, for example, girls who played musical instruments were weaker in mathematics than those who did not. Others did not recognize the need for a good *operational definition* of "good at math." Their questions merely reiterated the hypotheses: "Are you good at math?" Subsequently, when they attempted to display their data they found that literal reiteration of the hypotheses could not lay the groundwork for objective measurement of that relative term.

Not all the students who framed the essential questions stopped there, however. One way or another 69 percent at Grade 8 and 31 percent at Grade 12 went overboard with questions such as the following:

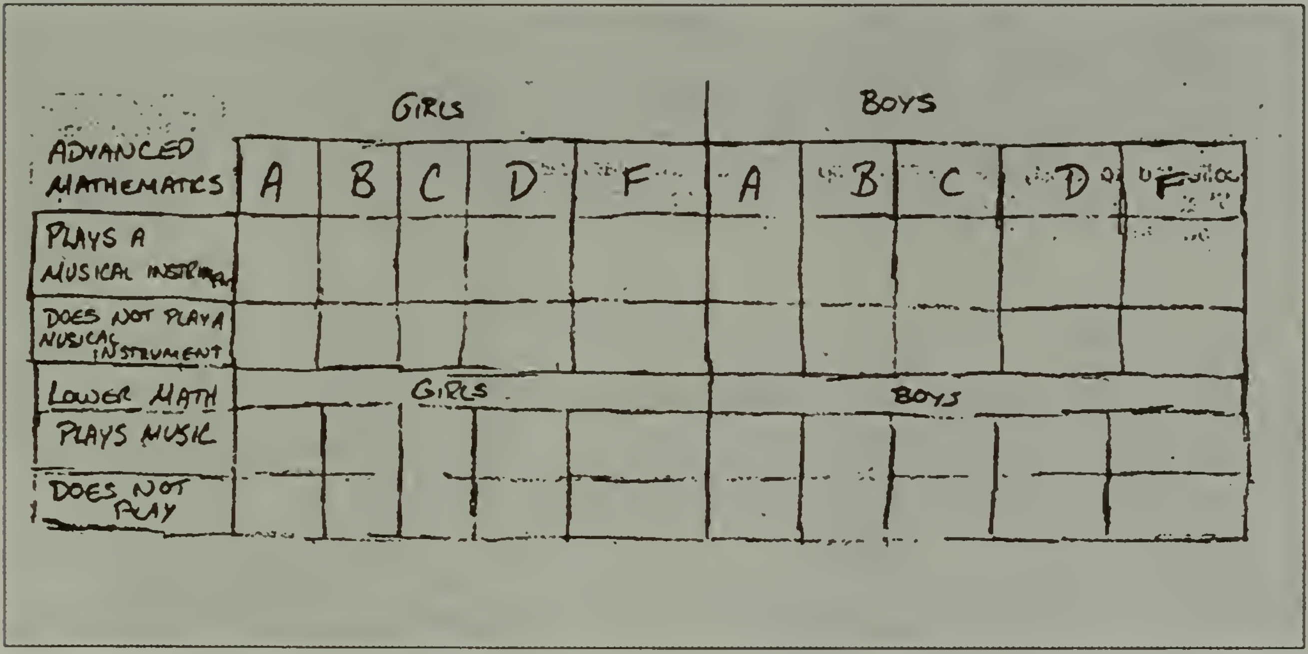
> Do you like math more than you like playing your musical instruments?
>
> Do you find math: (a) boring; (b) useless; (c) interesting; (d) fun?
>
> Boys: Do you know anyone who plays a musical instrument? If so, are they good in math? If they don't, are they good in math?

To put themselves in Wanda's shoes, however, students had to go beyond relevant questions and a "general" understanding of the term *hypothesis*. To envision proper analysis of data compiled through their questions, students were obliged to recognize that they were dealing with two distinct hypotheses, one the converse of the other. Equally important, each hypothesis involved two different types of variables. Whether or not students played a musical instrument was a categorical variable (usually students were asked whether or not they played). In contrast, math achievement was usually interpreted as linear (grades or test scores). In order to show the relationship between the two, math achievement had to be converted into a categorical form (pass/fail) or aggregated (mean or median grade).

Given the difficulty and complexity of asking the essential questions, formulating hypotheses distinctly, and coordinating salient variables, it is not surprising that only a quarter of the Grade 8 students and a third of the Grade 12 students were able to show clearly how results should be displayed. Of those, most used two bar charts to represent the difference in achievement for each subset of boys and girls.



A few showed or described other acceptable methods of displaying the data.



While the second example shows less data reduction, the pattern of values in the different cells would help to reveal the relationship being investigated. It is interesting that the student recognized the relativity of the term "good at math" by separating groups of students according to their level of mathematics courses.

Among eighth graders 50 percent and among twelfth graders 34 percent provided poor or irrelevant formats. Grade 8 students, in particular, hindered themselves with too many variables. Some focused on individuals, which meant that their results could only be displayed in chart form. They did not seem to understand that the data had to be reduced and organized before it could be visually presented.

2. Explain how Wanda might analyze and display her results. (For example, how could she set up a table or graph to help her test her hypotheses?)

Musical Instorments and Math

Student	math average	time on M. HW	time on instr	math
Anne	72%	20 minutes	1 hour	A
Lynn	56%	15 minutes	1/2 hour	B
Amy	90%	45 minutes	45 minutes	d
Sue	80%	1/2 hour	1 1/2 hours	B
Colleen	79%	1/2 hour	45 minutes	A
Mark	94%	1 hour	20 minutes	C
John	89%	45 minutes	20 minutes	d
Chris	87%	40 minutes	1/2 hour	D
mike	96%	1 1/2 hours	1 hour	C

PLEASE GO ON TO THE NEXT PAGE→

In addition to involving too many variables, the response above shows a raw data display and, therefore, fails to provide some means of analyzing and viewing results that would help Wanda test her hypotheses.

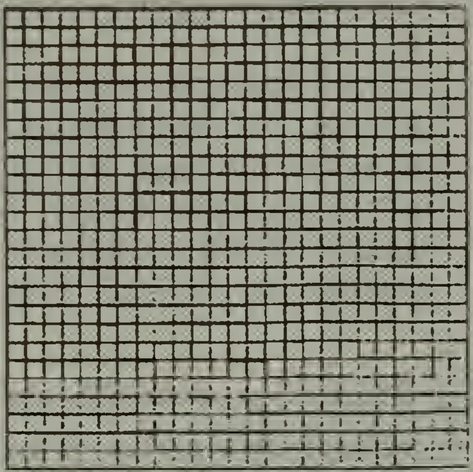
Perhaps more than any other question under Scientific Inquiry, this open-ended exercise revealed students' lack of conceptual clarity in designing experiments. Although nearly half in Grade 8 and more than half in Grade 12 survived the data-gathering phase, the next two steps – managing the data and displaying results – took a heavy toll. By the end of the exercise a large majority in both grades were frankly lost.

Brothers and Sisters: Grade Eight and Grade Twelve

Name	Sex	Age	Best Subject	Number of Brothers & Sisters	Grade in Reading
Adams	M	12	M	2	80
Archer	M	12	R	0	85
Carter	F	11	H	4	74
Carrera	F	12	R	0	87
Davenport	M	13	S	1	85
Davidson	F	12	M	1	77
Fenwick	M	12	M	0	79
Franklin	M	12	R	2	77
Harris	F	12	S	3	70
Jacobi	F	11	H	2	80
Kelley	M	12	S	0	83
LaFontaine	M	12	M	1	80
Miranda	F	12	H	1	76
Moore	F	13	S	1	82
Oberon	M	12	H	0	91
Peterson	M	12	M	1	86
Potvin	F	12	H	3	80
Smith	F	12	M	2	83
Trudeau	F	11	R	2	81
Washburn	F	12	R	4	72
Weinberg	M	13	S	3	75
Wilson	F	13	S	2	79

Sex: M = Male, F = Female
Best Subject: H = History, M = Math, R = Reading, S = Science

1. There are five groups of students based on the number of brothers and sisters they have. Compute an average reading score for each group. In the space at the right, make a graph showing your results.
2. What conclusion can you draw from your results?



Whereas in the previous exercise students were asked to identify the information to be collected and devise a way of displaying the results, the exercise shown here only required them to interpret data set before them. Grade 8 students, moreover, were told how to proceed with interpretation: group names by number of siblings, compute an average reading score for each group, graph the results on a prepared grid, and draw a conclusion.

Probably as a result of these explicit directions, Grade 8 respondents on balance dealt more competently with **Brothers and Sisters** than with **Wanda**. More than half stated a satisfactory conclusion. "People with no brothers or sisters have better grades," wrote a student who correctly calculated and plotted the averages. Some who drew the correct conclusion added plausible conjectures to their answer, as in the following example.

> People with no brothers got more help on homework from parents. But parents of many children have to split up their time.

Among Grade 8 students there was a strong correlation between arriving at the correct conclusion and graphing the data accurately. For purposes of assessment, however, group improvement between this exercise and the previous one was not the main message. Rather, what stood out was the large gap which separated those who could comprehend the task at all and those who could not.

A quarter of the students reached incorrect conclusions by desperate means.

> "The people with more brothers and sisters do better than the ones with no brothers and sisters," said one student. He had formed three groups from the rosters: individuals with 0 siblings, with 2 siblings, and 4 siblings. Then he averaged only two scores from each group and drew a line graph connecting three points representing these averages.
>
> "My conclusion is that the highest is science and math," said another. This student had counted down in units of 5 lines from the top of the chart, thus producing 4 groups of 5 names and a fifth group with 2 names. Thereupon he calculated the average number of siblings (fractional, of course) in each of these arbitrary clusters but graphed nothing.

Finally, 21 percent of the Grade 8 respondents left this exercise blank. That high number, combined with the even greater number of wild efforts such as

those sketched above, suggests that students may have been nonplussed by the columns of figures contained in the question. Although the skills actually required to handle this exercise could be regarded as low level, to such students every chart may appear as a random collection of letters and numbers in a rectangular white space for unknown purposes.

At Grade 12, 67 percent of the students drew the correct conclusion for this exercise even though it was presented to them in a more challenging manner. For these older students the same task, based on the same **Brothers and Sisters** chart, was posed as an open-ended question. Instead of being directed to compute averages and plot results on a prepared grid, Grade 12 students were simply asked to "use the data" to draw a conclusion. It was up to them to identify and handle dependent and independent variables. About half of those who performed adequately volunteered a graph or chart in support, but those who did not proved equally adept at relating number of siblings to grade averages by other means – e.g., reporting average scores in narrative form.

Once again, however, the gap between those who could deal with such problems and those who had no idea how to go about it was wide. Several respondents in the latter group chose wrong variables. Perhaps because reading was recorded as their "Best Subject," five individuals listed on the chart were mistaken by a few students for the "five groups" about which they were supposed to draw a conclusion. Several respondents evidently did not know how to compute an average. About 20 percent left the exercise entirely blank. (See Appendix.)

Understanding Scientific Concepts

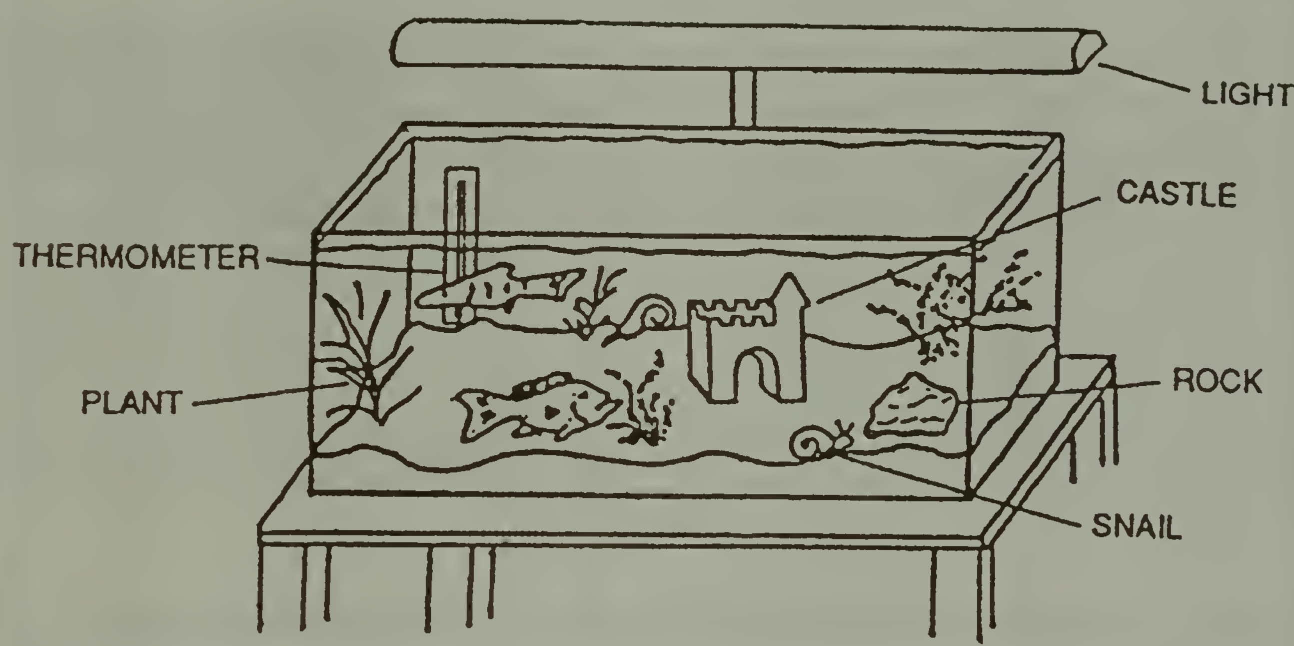
Obviously, science is more than inquiry. It is also a network of theories of how the world functions. An important aim of science education is to make students aware of such theories and help them to understand them. For this to occur, teachers must be aware of how their students rationalize the world about them. Approximately half of the open-ended questions were related to the content of science, examining how students understood certain everyday phenomena.

Life Science

Two questions were used to investigate students' understanding of concepts in life science. At the fourth grade students were asked about the ecosystem existing in an aquarium. Would they think of the objects in the aquarium as components of a livable environment for fish, or would they attribute some other, perhaps purely human meaning to these items?

At the eighth and twelfth grade levels, students were tested on their understanding of the transformation of energy that takes place in plant growth.

Aquarium: Grade Four



In the picture of an aquarium above, six items are labeled. Which of the six items are important to use in or with an aquarium? Explain why each one you name is important.

In general, students responded well to this exercise, focusing on conditions necessary for the fish to survive rather than features that would enhance their viewing. Almost all of them (96 percent) mentioned the thermometer as an indicator of water temperature, and some went on to explain the importance of regulating temperature for the good of the fish. Almost 80 percent named the light as an important feature. Most stated that its purpose was to enable the fish to see; others recognized its importance in plant growth; others suggested that the light was important mainly for the viewer. A sizeable number (20 percent) stated that the plant provided oxygen for the fish.

Among the 44 percent who cited the importance of the snail, approximately half offered a scientific justification such as keeping the tank clean and eating algae. The other half tended to anthropomorphize. "The snail would be a companion for the fish," said one.

Conceivably the castle and the rock could serve an environmental function – as places for fish to hide, for example, or as boundaries for territorial

species. However, the majority of students who cited these objects as important (slightly more than one-third) gave other reasons. Some students wrote that the fish would play around the objects, and others classified them as decoration.

The popularity of aquariums in the classroom and at home may account for the level of success on this question. Most of the students seemed to realize that fish require a highly determined environment to survive, unlike cats and dogs, and that the components are related in an "ecosystematic" way.

Tree: Grades Eight and Twelve

A small tree is planted in a meadow. After 20 years it has grown into a big tree, weighing 250 kg more than when it was planted.

Where do the extra 250 kg come from? Explain your answer as fully as you can.

Do students understand the basic processes involved in plant growth? There is evidence that younger students (Grade 6 and below) believe that plants do not make their own food, that a plant's food is absorbed from outside sources, and that energy is unimportant in plant growth. As children get older, they begin to understand how plants made food but not how food and energy are related in plants. This was the question that this item set out to explore. How well do older students understand these questions?

The item we used was developed and analyzed by the Assessment of Performance Unit (APU), a British assessment effort that functions much like the National Assessment of Educational Progress (NAEP). Despite its simplicity, it yielded useful information about how students understand plant growth and the function of photosynthesis. British reports suggest that this is poorly understood by students,[4] and results from the open-ended question confirm their findings.

At both grade levels, only about 10 percent of the students acknowledged the transformation of substance that is involved in plant growth. Half the students at both grade levels answered the question with some form of tautology, either stating that the extra mass comes from the growth of the tree (i.e., interpreting the question as "which *part* of the tree accounts for its extra weight?" or "what makes the tree heavy?" (See Appendix.)

4 Bell, B.F., and Brook, A. *Aspects of Secondary Students' Understanding of Plant Nutrition.* Children's Learning in Science Project. Leeds University, 1984.

Typical of such responses are:

> The 250 kg comes from the size of the tree. It has grown to be much larger. Now there are roots underground that also make it weigh more. The size and roots add a lot of weight to a tree. (Grade 12)

> As it grows older, the tree gains weight because it becomes larger and it can store more food and water. This holds true for a sponge. When it is dry, it's very light. As you add water, it comes heavier. Also, the leaves of the tree hold water brought up by the roots through the xylem tubes. The leaves also have weight that can make the tree heavier. (Grade 8)

Approximately 15 percent of the students referred to nutrients in the soil and, occasionally, to sunlight but did not discuss any transformation. The impression gathered from these answers was that students believed that everything that trees take in is retained as mass. Even among the students who did refer to "photosynthesis," many seemed to regard it as a product rather than a process."Photosynthesis is produced from the sun and the water and the soil," stated one student.

A further 15 percent of students appeared to believe that the extra weight was due to the absorption of water by the roots.

> The 250 kg comes mostly from water and the sun. As a tree grows it moves to the sun. Water in the form of rain is absorbed by the roots through the soil. This is what makes the tree grow. (Grade 8)

The role of tree roots in plant growth is commonly misconceived by students. At Grade 4, the following responses were given to a multiple-choice question.

> What part of a plant takes in sunlight and makes food for a plant?

41%		roots
41	*	leaves
9		petals
9		flowers

In a more directed question at Grade 8, a quarter of the students indicated their belief that photosynthesis takes place in the roots of a plant.

Although the question was broadly stated, perhaps encouraging a generalized and simplistic response, that so many students at both grades did not appear to recognize any transformational process in plant growth is disturbing. It casts doubt on the level of understanding that they have of ecosystems in general and photosynthesis in particular.

This was suggested by the following multiple-choice items in the assessment:

What are two products produced by plants during photosynthesis?

Gr 8	Gr 12		
16%	13%		sugar and carbon dioxide
27	29	*	sugar and oxygen
22	22		water and carbon dioxide
26	32		water and oxygen
8	4		sugar and water

Photosynthesis is the way plants make food. Which of the following are always needed for photosynthesis?

Gr8	Gr12		
51%	55%	*	water, light, chlorophyll, carbon dioxide
32	32		water, light, chlorophyll, oxygen
6	5		water, roots, stems, leaves
10	7		water, light, fertilizer, air

There are two obvious conclusions from these results: (1) a sizeable proportion of students do not know the basic facts concerning photosynthesis (its requirements and products); and (2) that there is almost no growth in knowledge beyond the eighth grade level.

Although some (from a quarter to a half) appear to be able to give the chemical results and requirements of the process when given the term itself (see multiple-choice above), many fewer are able to recognize it within the context of the environment. When photosynthesis is taught and tested with reference to a learned equation, students appear to give a kind of conditioned response, without relating what they have learned to the world about them. Most students do not seem to understand photosynthesis as a process involving energy transformation and chemical reactions. Instead, they seem to

believe that everything that a tree takes in is retained as mass. They appeared to be simplistic in their approach (e.g., "You water a house plant and it grows"), without appreciating the depth of the processes involved.

Physical Science

Students' understanding of physical science was investigated at all three grade levels. At Grade 4, the question centered on students' understanding of the nature of material. At Grade 8 and Grade 12, students were asked about the concepts of energy and electrical conductivity.

Materials: Grade Four

Objects are often made of more than one material. What are two materials used to make each of the items listed below? *Why* is each material used?

<u>bicycle wheel</u>

Material 1: Why?

Material 2: Why?

<u>frying pan</u>

Material 1: Why?

Material 2: Why?

Approximately two-thirds of the students gave metal and almost 80 percent gave rubber as the two materials used in the manufacture of a bicycle. Far fewer (approximately 40 percent) were able to explain why each was chosen.

In response to the frying pan question, 83 percent of students named metal as one material and 62 percent mentioned plastic, wood, or rubber as the second. Approximately half gave an appropriate justification for the use of metal, while slightly fewer were able to justify the use of plastic, wood, or rubber for the handle.

The types of justification used for the materials differed. Students seemed to have a clear idea of the characteristics of metal. They referred to its strength when describing its use in the bicycle and seemed particularly aware of its ability to conduct heat (in the frying pan). When discussing the use of rubber

in the construction of the bicycle, the majority of students who answered correctly appeared to refer to their own experience. For example, they spoke of its ability to produce a smoother ride.

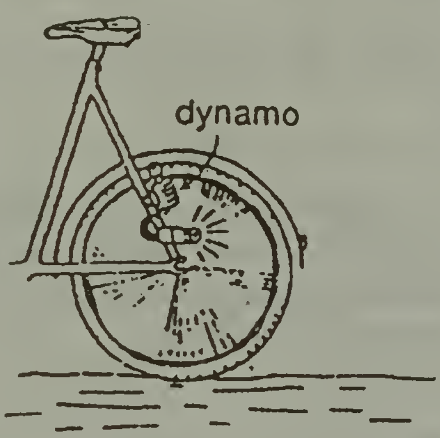
The role of practical experience was also seen in students' discussion of materials for the handle of a frying pan. Most of the students who answered correctly wrote in terms of the need to pick up the frying pan. In this situation, the characteristic of the material is less important than *why* it is being used. The characteristics of different materials are more likely to be the concern of school – in contrast to home – instruction. Finally, many showed good observational skills, referring to their own experience with cooking. The influence of prior knowledge on achievement is apparent in all content areas.

Dynamo: Grades Eight and Twelve

A girl has a dynamo on her bicycle to light her lamps. She notices that when the dynamo is being used it is harder to pedal at the same speed.

She is told that this is because energy cannot be created or destroyed, only changed from one form into another.

How does this explain what she has noticed?



Although students have many experiences in energy activities throughout their school years, the conservation of energy (the idea that the total amount of energy within a system remains constant during transfers and transformations within a system) is a difficult concept for them. Part of its difficulty lies with the different ways in which the word *energy* is used in a scientific context, as contrasted to its use as a metaphor in everyday language. In addition, the notion of a constant quantity is counter-intuitive to students when they are faced with what seems to be an increase or decrease of energy in natural phenomena. How well they understand the law of conservation and how they grow in that understanding is the focus of this question, which was taken from the APU.

When the question was presented in open-ended format, responses showed a large growth in sophistication between the eighth and twelfth grades. (See Appendix.) While 46 percent of eighth graders responded in terms of energy conservation, 74 percent of twelfth graders did so. Slightly more than a quarter of the eighth graders showed good reasoning, as illustrated below:

> She has noticed that it is harder to pedal because there is more energy being created into electricity. So she must pedal harder in order to keep the same speed.

On the other hand, a quarter of the students attributed the need for increased effort to the mechanical action involved, primarily friction.

> The dynamo causes friction, slowing down the wheel, causing it to be harder for the person pedalling. The same thing is true for having to swim upstream instead of across.

Another fifth of eighth graders clearly misunderstood the concept or the question.

> The dynamo isn't making her go faster, so it's not creating anything, and it's not making her go slower either, which means it hasn't destroyed anything. It's only making it hard to pedal when it used to be easier.

In contrast, approximately half the twelfth graders showed good reasoning ability in their responses.

> The dynamo takes the energy that she uses to pedal the bicycle. Without the dynamo she would only have to move the wheel. However, since she has to move the wheel *and* provide the dynamo with energy, more energy is required from the girl.

In the multiple-choice form, students were not given the definition of the law of conservation. Rather, they were asked to use it, with reference to the situation.

Maria has a generator light on her bicycle. She notices that when the light is being used it is harder to pedal at the same speed. How does the law of conservation of energy explain what Maria has noticed?

Gr 8	Gr 12		
33%	34%		The generator light rubs against the wheel and slows the bicycle.
13	9		When energy from pedalling is transferred to the light, some of the energy is lost to the atmosphere.
40	44	*	Some of the energy used to pedal the bicycle is changed into light and heat energy.
13	13		The amount of energy required to pedal at a certain speed is always the same; the light just makes it seem harder.

At Grade 8 results in response to the two forms were comparable: approximately the same number of students showed an awareness of the transfer of energy from the pedaling motion to the light. However, the results at the Grade 12 level differed substantially. When students were not explicitly given the law of conservation, there was little evidence of growth of understanding.

This casts doubt on the degree to which the open-ended responses of the twelfth graders reflected their understanding of the phenomenon. Although 25 percent more twelfth than eighth graders responded satisfactorily to the open-ended version of this question, their responses may have reflected their increased ability to interpret the given statement appropriately rather than an increased understanding of the concept involved. In other words, their increased verbal and reasoning ability may have accounted for their success.

Fuse: Grades Eight and Twelve

Explain in detail how a fuse works in an electrical circuit.

When this question was presented in multiple-choice form on the assessment, approximately half the twelfth grade students chose "to stop the current when the wires overheat" as the correct option. However, when presented in an open-ended form, it appeared to be regarded as a difficult and unusual question by most students. Almost a third of the eighth grade and nearly that many twelfth grade students did not attempt it. Only 12 percent of the eighth graders gave a satisfactory response. (See Appendix.)

At Grade 12 success rose to 21 percent, with half of these students giving a technically correct response, the other half replying in a non-technical way. Another 10 percent at both grade levels replied by stating a general purpose of a fuse (e.g., to control the flow of electricity) but did not discuss how it accomplishes its purpose.

The other responses were divided between those who maintained that a fuse generates electricity, that a fuse stores or releases electricity, and that it measures electricity.

> The fuse is set up so that it may travel to whatever appliance or machine is needed. The fuse stores electrical energy which then travels to the circuit where the appliance is plugged into, and the energy travels into the appliance. When a fuse goes dead, or "blows," we replace it with a new one which will do the same thing, only with more electrical power. (Grade 8)

> Fuses are used to regulate voltage in an electric circuit. If it were not for fuses, lightbulbs would pop at the flip of a switch. The amount of electricity would damage all household appliances. Uncontrollable fires could break out if the conditions were right. People shouldn't complain when they "blow a fuse" and have to replace it. (Grade 8)

These answers are consistent with those given in response to a similar multiple-choice item. Although, when presented with a set of preselected responses, 49 percent of students chose the correct answer, the others chose responses resembling those above.

Gender differences were apparent in responses to this question, particularly at the Grade 12 level. While 32 percent of males gave a correct answer, only 9 percent of females did so. While almost 30 percent of females stated that the purpose of a fuse was to generate, store, release, or measure electricity, almost no males answered in this fashion. Furthermore, at both levels, between 35 and 40 percent of the females did not even attempt to answer the question. On the other hand, at Grade 12 more males gave an irrelevant answer or one that we termed as on-topic gibberish.

> Seeing how my last science course was chemistry and since I only took a half year of physics, I cannot explain this problem in detail. I believe that the flow of oppositely charged protons and electrons is what makes a fuse work. (Grade 12)

This discrepancy in responses between the genders confirms a general finding in other research and other assessment programs. As observed by the authors of a recent report on science achievement as measured by the National Assessment for Educational Progress:

> A large gender gap existed for students who had not taken Physics and, although course-taking appeared to improve performance substantially for both males and females in this content area, it did nothing to reduce – and in fact may have increased – the difference in performance between them.[5]

Undoubtedly, different experiences and interests may also have influenced the results obtained by girls and boys. The NAEP reports the following percentages of students responding "many times" to the question, "Have you ever tried to fix something electrical?"

	males	females
Grade 7	36	10
Grade 11	46	8

We can only speculate that the experience of students in Massachusetts does not differ substantially from that of students nationally.

5 Mullis, I., and Jenkins, L. *The Science Report Card.* Princeton, N.J.: Educational Testing Service, 1988.

Notes

Summary and Implications for Instruction

The open-ended questions allowed us the opportunity to look at two distinct aspects of science comprehension. One we have called scientific inquiry, the extent to which students are able to think and act like scientists. The other is in the area of scientific concepts, how students understand scientists' explanations of the world.

In order to measure the extent to which students act like scientists, we presented some problems to them and asked: What can you do? (*Not* what are the rules?) Taken as a whole, the results from these open-ended tasks, which required the application of scientific procedures – in contrast to the verbal knowledge of scientific procedures – were disappointing, particularly in light of students' results on multiple-choice questions. Although students appear to know and recognize the rules and principles of scientific inquiry when presented as stated options, unstructured situations that demand an application of these principles seem to baffle them. Asking them to *do* rather than to *recognize*, made apparent their lack of actual understanding. With this consideration in mind, let us go back to the test questions.

Spot Remover: Reflecting the less formal nature of science education at the primary grades, this report contains only one Grade 4 question in the area of Scientific Inquiry. **Spot Remover** required a recognition of different factors that might be controlled in an experiment. Although a knowledge of the control of variables is not something that is formally tested at this level, some general notions of its importance are assumed in different questions on the multiple-choice tests. On the open-ended, more than half of the fourth graders either did not understand the question or were not able to list any factors that should be kept constant in an experiment. This suggests that

younger students may not understand what it means to "do an experiment," since the control of variables is a fundamental to any experimental design.

Although older students were more successful with a similar but more open-ended question, which required them to recognize the requirements of the task for themselves, there remained a large discrepancy between the percentage of students who are able to recognize a "rule," such as "control of variables is essential in an experiment," and those who are able to recognize when and how such a rule should be applied. Multiple-choice questions do not adequately measure this ability. Specifically, despite the fact that over 80 percent of twelfth grade students can recognize the need for controls when presented with a set of options, when required to generate ideas, such as outlining an experiment themselves, they exhibit uncertainty about the principles involved.

Homework. This question was given only to twelfth graders. Again, it dealt with the control of variables in an experiment; yet this question required students not only to recognize the experimental variable but to evaluate an experimental design. Only a third were able to tackle the task successfully. Their familiarity with the context appeared to make it more difficult for them to remain objective and to assume a scientific posture — confining their remarks to the design of the experiment rather than commenting on the substance (i.e., the efficacy of homework). Again, students' success in completing this task was lower than when they were asked to critique experiments in multiple-choice format; although even here only half of the students tested were able to choose the correct options.

Although difficult, this type of question exemplifies an important aspect of scientific thinking, the ability to analyze and critique experiments. In contrast to devising an experiment, the analysis requires the kind of thinking that is particularly relevant to students' lives as consumers of information. Not everything that is labeled "scientific proof" is, indeed, valid. Students should be capable of examining such proof by asking questions. "What is the basis for the claims? Are the results founded on good evidence? Are the conclusions justified by the results?" This is a kind of thinking that needs both practice and discussion.

Wanda. This question tested both concepts and procedures at the eighth and twelfth grade levels and revealed shaky understandings of the meaning of

hypothesis and the need for an operational definition of a relative term. Asked to design a questionnaire that would test two hypotheses: that girls who play a musical instrument are not as good in math as those who don't and the opposite for boys, many students appeared to believe that these hypotheses were givens and interpreted the task as one of establishing *why* this was the case. Others merely reiterated the term, "good at math," without recognizing the need for quantifying it. As a result, only 42 percent of eighth graders and 58 percent of twelfth graders were able to generate the essential thinking required. However, students found the greatest difficulty in organizing and reducing the data. Only a quarter of the Grade 8 students and a third of Grade 12 students succeeded in displaying the results in an interpretable fashion — either in graph or table format.

It is generally agreed that, although almost all students can read graphs and most can interpret them, they experience difficulty when asked to make their own displays of data. Consequently, instruction usually focuses on graphing skills, such as plotting, ordering, and scaling. What is less acknowledged is the need to understand the kind of variables that are being displayed and how they relate to the question asked. In order to encourage this understanding, students should be given more opportunity to generate their own data, to discuss the kinds of variables involved, and to relate these to the display.

Brothers and Sisters. The final question under the Scientific Inquiry category explored students' ability to interpret data that was presented in tabular form. Grade 8 students were told how to proceed with the interpretation, while Grade 12 students were given more open-ended directions to "use the data." More than half of the eighth graders came to a correct conclusion in this — perhaps more familiar — task; however, the other half either did not even attempt to answer or showed that they had little idea of how to proceed. Two-thirds of the twelfth graders completed the task successfully. Again, students showed the need to practice in reducing data and discovering patterns that occur in a series of changes.

The second aspect of science that we examined was students' conceptions of some basic scientific theorems: the interdependence of nature, plant growth, characteristics of materials, the conservation of energy, electrical conductivity. The questions were simple, within a context of everyday experience. They did not explicitly ask for the more formal, technical aspects of theories. Rather, they examined students' ideas of phenomena. As in other studies, we found widespread misconceptions. More surprising, however, we found a

resistance to such kinds of questions, suggesting that students perceived this type of questioning to be unfamiliar and threatening.

Aquarium. This question, given at the fourth grade, asked about the ecosystem that exists in an aquarium. Students were asked to explain why different components of an aquarium might be important to fish life, and the exercise proved to be provocative and interesting. Most children answered fully and seemed to show a good knowledge of the functions of different elements. We suspect that children's observations of aquariums in the classroom, where discussions might arise, contributed to their understanding of the interdependency that exists within such a closed system.

Tree. Eighth and twelfth graders experienced more difficulty in accounting for the growth of a tree. The open-ended nature of the question ("Where does the extra weight come from?") may have led half the students in both grades to reinterpret the question as "Which part of the tree accounts for its extra weight?" or "What makes the tree heavier?" At the same time, among those who appeared to understand the requirements of the question, few referred to the transformational process in plant growth or the manufacture of new cells. The majority appeared to believe that growth is a process of absorption. That results were similar for the two grades suggests that this aspect of life science is not a part of the high school curriculum and that early misconceptions continue to exist throughout the school years.

Materials. As in the case of the aquarium question, children's practical experiences were apparent in their discussions of why specific materials would be chosen in the manufacture of different items. Although the large majority of fourth graders correctly identified the appropriate materials, far fewer were able to justify their answers. Those who did tended to refer to their own experiences with either riding bicycles or picking up frying pans. However, many also referred to strength as a property of metal — a more abstract justification than that based on personal experience.

It is obvious that students bring to school knowledge gained from experience in the world. School instruction should attempt to build on these experiences in order to broaden children's understanding. Discussions that relate the characteristics of different materials to students' observations of how and why

those materials are used allows them to connect their own knowledge with the more structured classification that takes place in the class.

Dynamo. Approximately half the eighth graders and three-quarters of twelfth graders responded satisfactorily to this question concerning the application of the law of conservation of energy. This result was in sharp contrast to results on a similar multiple-choice item, which asked the same question without explicitly stating the law.

Fuse. This final question in the area of physical science asked eighth grade students and twelfth grade students to explain the working of a fuse in an electrical circuit. Most students saw this as an unusual and difficult task, and some expressed their indignation at being asked such a question. A large proportion did not even attempt to answer. Only 12 percent of eighth graders and 21 percent of twelfth graders gave a satisfactory explanation. We see once more that the open-ended nature of the question may have disconcerted many students. When the question appeared in a multiple-choice format, approximately half of the twelfth graders chose the correct option. However, the open-ended task clearly showed their inability to generate a satisfactory answer. This question was also one that showed a clear gender gap. Many fewer girls than boys attempted the question, particularly at the older level where 40 percent of the girls (in contrast to 16 percent of boys) left the question blank.

Upon examining such test results, Mary Ann Sudolnik, a practicing teacher and longtime member of our Science Advisory Committee, offered several acute observations:

> It appears that there is a very large gap in our children's scientific education. It is in the application of the scientific method to real problems. Students were unable to identify variables, gather information, and then reduce the information into something that would lead to a reasonable conclusion. The instructional implication is that what we present to the student is too cut-and-dried. They need a lot more experience with these areas. As teachers, we must make them aware by having them design and follow through on all the steps of a true scientific investigation.
>
> Conceptual understanding of scientific principles should be evaluated frequently enough so that the teacher has a better understanding of what the student really has internalized about the information.

> We need to do science, not read science, not memorize science but actively engage students in get-our-hands-dirty kind of science. This hands-on approach would go a long way towards eliminating the problem area that I saw in this assessment testing.

The student difficulties referred to in this statement are related to the uncertainty science educators feel about the rationale for laboratory activities. Is the purpose of such activities to demonstrate scientific concepts or to perform authentic investigations? In the typically structured lab experience, variables are controlled for the student. He or she may conceive the "problem" to be: have I gotten the right results? Although such activities may be useful in illustrating natural phenomena and set procedures, their usefulness in developing a scientific approach and understanding that can be generalized to the students' own lives has been questioned.[6] Perhaps allowing students to resolve inconsistencies, to generate their own experiments, and to critique the experiments of others might increase their understanding of both concepts and methods by which those concepts can be tested.

Although scientific inquiry is often regarded as "process" and scientific knowledge as "understanding," the two must progress together. After all, theories about life and physical functions are based on evidence gathered through the process of scientific inquiry. They are not final, but are tentative explanations based on data and subject to revision. Unfortunately, schools often deprive students of an understanding of this aspect of the scientific enterprise, treating all science as given.

Students' responses to the set of questions reported here underlines the urgency of a general concern in science education — that our assumptions of what students know and understand are unfounded. Once students have been asked to explain or apply concepts, it is apparent that their own intuitive understandings exert a far great force in explanation than do those of texts and teachers. Although students can be made aware of accepted scientific theories through teaching, it is generally believed that understanding is a more complex process, requiring the examination and testing of more intuitive — and often incorrect — notions of how the world functions. For the latter to occur, teachers must be aware of how their students rationalize the world about them. Open-ended questioning appears to lead in the right direction.

6 Bates, G. "The Role of the Laboratory in Secondary School Science Programs," in Rowe. M (ed.), *What Research Says to the Teacher*, Washington, D.C.: NSTA Publications, 1978.

Appendix

Spot Remover

Reporting Category: Scientific Inquiry
Grade Level: Fourth

Number of Correct Variables (percentage of responses): *

3 or more	13
2	17
1	13
0	42
blank	15

Types of Variables Cited (percentage of responses):

type of stain	25
type of fabric	20
amount of spot remover	12
amount of stain	10
washing time	10

Examples:

Appropriate:

The stain should be the same, the material should be the same, he should give each spot remover the same time to remove the spot.

Inappropriate:

Spot remover B will probably be the same color, or smell the same, or have some of the same stuff in it, or it might cost the same or be from the same store.

Both spot removers will at least get out a little stain.

* Throughout the tables, percentages may not total 100% because of rounding.

Grade Level: Eight and Twelve

Response Summary:	Frequency (%)	
	Gr 8	Gr 12
Controls for 1 or more variables in addition to stain and fabric and replicates for different stains and fabrics	8	42
Controls for one or more variables in addition to stain and fabric, no replication	30	15
Controls for stain and fabric only	27	16
No control	28	20
Blank/Irrelevant	7	5

Homework

Reporting Category: Scientific Inquiry
Grade Level: Twelve

Response Summary:	Frequency (%)
a. Identified problem with controls	20
b. Identified other concerns, e.g., size of sample, replication	16
c. General opinions about homework	25
d. "poor experiment"; poor reasons given	8
e. "good experiment"	14
g. Blank/Irrelevant	18

Examples:

a. I think the teacher's conclusion was incorrect. Even though the first group did not have assignments to be completed at home, they did get extra practice with the added worksheets. Without these, they probably would have performed differently on the test.

b. His test sample is too small to get any accurate results to begin with. There are also too many substitute factors, such as worksheets and shortened lab time which make the classes unequal. To improve the test the teacher needs more tests with a greater number of students when all conditions but the doing of homework are the same.

c. It's a good one because homework makes no difference. I never get a lot out of homework. I would get more out of it doing it at school.

d. No, it wasn't a good one because the students that didn't have homework could still study for the test. The teacher should have told the students they couldn't study.

e. The conclusion was a good one but I feel that it all depends on how much the students are willing to study.

Brothers and Sisters

Reporting Category: Scientific Inquiry
Grade Level: Eighth

Response Categories

correct conclusion, good analysis:

People with no brothers got more help on homework from parents. But parents of many children have to split up their time. So neither of the children get full help.

(Student averaged scores from each category and drew a correctly labeled bar graph.)

If you have many brothers and sisters, you'll be distracted more often. If you are distracted more often, you won't get as much quality reading time. If you don't get enough quality reading time, your reading grades will be low.

(Student averaged each group and plotted results.)

correct conclusion, poor analysis:

The people with more brothers and sisters do better than the ones with no brothers and sisters.

(Student averaged first two scores in the following groups: no siblings, 2 siblings, 4 siblings. He next drew a line graph connecting the 3 points.)

incorrect conclusion, poor analysis:

The less kids are home (brothers and sisters), the lower the grade gets on most of them. The total average of the 5 is 80.4 avg.

(Student averaged the scores of 5 children, 2 of whom had no siblings, 2 had 2 siblings, 1 had 4 siblings. Conclusion was from inspection of these cases. No attempt at graphing.)

6 people get about 70 in reading.
11 people get about 80 in reading.
1 person got a 90 in reading.
Most people got a 80 average in reading.

(Student tallied names under columns headed "70, 80, 90." Conclusions were based on visual inspection of the tallies. This was accompanied by a bar graph, correctly drawn and labeled "M" "F.")

My conclusion is that the highest is science and math.

(No computation or graph.)

Group 1 (1.4) B and S

Group 2 (1.6) B and S

.

.

.

(Student divided the list into 4 groups of 5 with a fifth group of 2 and calculated the average number of siblings for each group.)

Tree

Reporting Category: Life Science
Grade Level: Eighth and Twelfth

Response Summary:	Frequency (%) Gr 8	Gr 12
a. Tree makes tissue from constituents it takes from the environment	12	7
b. Tree grows because of what it takes in from its environment (no indication of process)	33	30
c. Tautology	48	50
d. Irrelevant	3	7
e. Blank	4	7

Examples:

a. The other 250 kg comes from nutrients in the soil and air. Through photosynthesis the tree produces sugar from carbon dioxide in the air and water from the ground. The cells convert the sugar to other compounds. The other essential nutrients, such as nitrogen are also gained from the soil. Nitrogen is essential in all stages of tree development and living. (Grade 12)

a. When the small seedling is planted it is nourished with water and sunlight, just like the way we humans grow. This is the tree's energy, like food is our energy. The tree changes the water it receives, by sunlight and itself, into energy, which makes the tree grow. This process is called photosynthesis. The tree grows and uses the energy to become larger in size. So, the water and sunlight is all the tree really needs to live and grow. (Grade 8)

b. The extra mass comes from the nutrients which the tree has taken in for the last 20 years. Nutrients such as water make the tree grow and photosynthesis and other processes mature the leaves and trunk. (Grade 12)

b. The tree has gained weight from the nutrients in the soil and has used those by mixing them with water and the rays from the sun. (Grade 12)

b. The extra mass was gained through the absorption of minerals that are water soluble which were absorbed through the roots. (Grade 12)

c. It came from all the sunlight, water and minerals in the ground. Which caused the roots to get bigger expanding the tree and making it gain weight. (Grade 8)

c. When you first plant a tree there is just a little bit of cells in it. Over the years when the tree is enlarging the cells are also splitting and enlarging. Therefore making more stronger bigger cells. By the time the tree is full grown all of those cells made the tree grow and become very heavy. (Grade 8)

Dynamo

Reporting Category: Physical Science
Grade Level: Eighth and Twelfth

Response Summary:	Frequency (%) Gr 8	Gr 12
a. Reasoning in terms of energy conservation	46	74
b. Description of mechanical action (e.g., dynamo rubs against wheel and slows bike)	23	11
c. Misunderstands concept or question	21	5
d. Blank	10	10

Examples:

a. The dynamo takes the energy that she uses to pedal the bicycle. Without the dynamo she would only have to move the wheel. However, since she has to move the wheel *and* provide the dynamo with energy, more energy is required from the girl. (Grade 12)

a. She has noticed that it is harder to pedal because there is more energy being created into electricity. So she must pedal harder in order to keep the same speed. (Grade 8)

b. Energy that she gives the bicycle is used to turn the wheel. When the dynamo is being used, it adds resistance to the wheel and uses some of the energy she gave to the wheel. Since energy is not created or destroyed she has to push harder on the pedal to move the wheel the same speed. (Grade 12).

b. When she didn't have the dynamo it was pretty easy to pedal with, but when she put on the dynamo it was harder because an object (dynamo) was put on and was rubbing against something so it made it difficult. (Grade 8)

c. The dynamo isn't making her go faster, so it's not creating anything, and it's not making her go slower either, which means it hasn't destroyed anything. It's only making it hard to pedal when it used to be easier. (Grade 8)

Fuse

Reporting Category: Physical Science
Grade Level: Eighth and Twelfth

Response Summary:

	Frequency (%)			
	Gr 8		Gr 12	
	M	F	M	F
a. Technically correct	2	2	18	3
b. Nontechnically correct	10	11	14	6
c. What fuse does but not how	15	4	9	8
d. General idea but some confusion	10	6	5	4
e. Incorrect	36	42	37	41
f. Blank	27	35	16	40

Examples:

a. A fuse is a thin piece of conducting metal that relays electrical energy. When the energy becomes too great the metal of the fuse breaks, breaking the circuit. The purpose of this is to help regulate power surges on appliances which could be damaged by power surges. (Grade 12)

b. A fuse has the same concept as a circuit breaker in that if there is an over-load or a short in the circuit the fuse will break inside and will stop the flow of electricity. (Grade 12)

c. A fuse stops too much electricity from going into a circuit. If too much electricity goes into the fuse it burns out and stops the flow of electricity. (Grade 12)

d. A fuse in an electrical circuit is a safety device made of a low-melting point metal alloy. When electricity flows throughout the circuit and the temperature gets too high, the fuse melts and opens another circuit to handle the overflow of electricity. (Grade 12)

e. A fuse in an electrical circuit lights up when it receives charges from both positive and negative poles. Without it the circuit would not be complete and the fuse acts as a conductor in the circuit. (Grade 8)

MASSACHUSETTS DEPARTMENT OF EDUCATION

REGIONAL CENTERS

Marcy Devanney
Northwest Regional Center
Berkshire Plaza
37 Main Street
North Adams, MA 01248
413/664-7101

Paul Burnim
Greater Springfield Regional Center
Macek Drive
Chicopee, MA 01013
413/594-8511

Amy Sosman
Northeast Regional Center
Jefferson Office Park
790 Turnpike St.
North Andover, MA 01845
508/689-0164 or 617/727-0600

Laurie Slobody
Central Mass. Regional Center
Beaman Street, Route 140
West Boylston, MA 01583
508/835-6266

Peter Cirioni
Greater Boston Regional Center
75 Acton Street
Arlington, MA 02174
617/641-4870

Pat O'Brien
Southeast Regional Center
P.O. Box 29
Middleboro, MA 02346
508/947-1231

Related publications about the Massachusetts Educational Assessment Program:

- *The Massachusetts Educational Assessment Program: 1988 Statewide Summary*
- *Using the School Report*
- *Reading and Thinking: A New Framework for Comprehension*
- *Science in the Elementary Schools*
- *Moving Geometry from the Back of the Book: A Report on Geometry and Measurement in the 1986 Assessment*
- *Social Studies Report*
- *On Their Own: Student Response to Open-Ended Tests in Reading*
- *On Their Own: Student Response to Open-Ended Tests in Mathematics*
- *On Their Own: Student Response to Open-Ended Tests in Science*
- *On Their Own: Student Response to Open-Ended Tests in Social Studies*

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST CLASS PERMIT NO. 102 DOVER, NH

POSTAGE WILL BE PAID BY ADDRESSEE

Advanced Systems
IN MEASUREMENT & EVALUATION, INC.
P.O. BOX 1217
DOVER, NEW HAMPSHIRE 03820-9979

APPLICATION

Massachusetts Educational Assessment Program
Open-Ended Scoring Session

Teachers and coordinators who have been involved in the scoring of the open-ended questions have found it to be a worthwhile and interesting professional experience. This year, the reading and scoring of the 1990 open-ended questions will take place at a two-day session during the week of June 25-29. If you would like to participate, please fill in and mail this application and you will be sent further details.

Name ______________________________

Home Address ______________________________

____

Home Telephone ______________________________

School Name ______________________________

School Address ______________________________

____

School Telephone ______________________________

I would be interested in scoring the following grades and subject areas:

GRADE LEVEL:

- ☐ Fourth
- ☐ Eighth
- ☐ Twelfth

SUBJECT AREA:

- ☐ Reading
- ☐ Mathematics
- ☐ Science
- ☐ Social Studies

designed by

Graphitopia